

DOI: 10.7652/xjtuxb201912022

管道天然气多组分浓度在线监测方法

敖家佩¹, 邱海峰², 贾唐浩¹, 李昕¹, 刘卫华¹, 雷绍充¹

(1. 西安交通大学电子与信息学部, 710049, 西安; 2. 中国石化西北油田分公司, 830000, 乌鲁木齐)

摘要: 为了解决管道内天然气监测技术中由于各组分互相交叉而带来的组分与浓度识别精度低且无法在线监测的难题,提出了一种采用气体传感器阵列与信息融合技术相结合的多组分浓度在线监测方法。该方法首先通过传感器阵列获得对混合气体的交叉干扰响应值,然后利用比较结果最优的辨识回归算法 MLP 对测量的实验样本进行组分辨识后对交叉干扰量进行修正并融合,得到相关权重与偏置参数,最终经过非线性计算获得满足精度要求的浓度值。仿真结果表明:采用该方法对开源数据库中 6 种单一组分进行种类识别,整体分类准确性最高达到 99%,对浓度进行回归识别中,回归系数 R^2 最高达 0.99。将该方法用于管道内天然气混合气体的种类与浓度识别,可得到甲烷、非甲烷总烃、氮气的回归相关系数 R^2 分别为 0.99、0.96 和 0.99。实验结果表明,该方法降低了交叉敏感的影响,满足了设备所需的精度要求,可以实现天然气的在线监测。

关键词: 在线监测;管道天然气;组分;浓度

中图分类号: TP277 **文献标志码:** A **文章编号:** 0253-987X(2019)12-0170-07

A Real-Time Monitoring Method of Components and Concentrations of Pipeline Natural Gas

AO Jiawei¹, QIU Haifeng², JIA Tanghao¹, LI Xin¹, LIU Weihua¹, LEI Shaochong¹

(1. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. Sinopec Northwest Oilfield Branch, Urumchi 830000, China)

Abstract: An on-line monitoring method of components and concentrations of pipeline natural gas is proposed to solve the problems that the recognition precision of components and concentrations in pipeline natural gas is low and the chromatography cannot be used online due to the cross interference of gas sensor array. This method combines a gas sensor array and information fusion technology. At first, the cross-interference response values of mixed gases are obtained through the sensor array, and the existing identification and regression algorithm MLP is used to identify the component gases and to correct the cross-interference quantity. Then the correlation weights and offset parameters are obtained. Finally, the concentration values satisfying precision requirements are obtained through nonlinear calculation. Simulation results of identifying six individual components in the open source database show that the overall classification accuracy of the proposed method is up to 99%, and the regression coefficient R^2 is up to 0.99 in the regression identification of the concentrations. When the method is applied to the identification of the components and concentrations of natural gas in a pipeline, the values of R^2 of methane, non-

收稿日期: 2019-07-02。 作者简介: 敖家佩(1995—),女,硕士生;李昕(通信作者),女,教授。 基金项目: 国家自然科学基金资助项目(51625504,61671368)。

网络出版时间: 2019-10-09

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20191008.2005.004.html>

methane hydrocarbon and nitrogen are 0.99, 0.96 and 0.99, respectively. These results show that the method reduces the influence of cross sensitivity, meets the accuracy requirements of equipment, and realizes on-line monitoring of natural gases.

Keywords: on-line monitoring; pipeline natural gas; component; concentration

随着西气东输工程投入运营,国内天然气行业进入了高速发展阶段。对管道内天然气成分和浓度进行定性定量分析能为及时改进产品质量提供保证。所以,生产单位急需在线监测设备,从而能够时刻监测到管内气体状态,观察气体开采的过程是否正常,开采气质量是否合格。由于气体计量参数多,管道距离长,采集管道内部压力较大,情况复杂,若按传统的设备维护工作方式,设备维护人员必须到达现场,并对设备进行深度诊断维护,这个过程本身具有一定的危险性,而且工作量大,很难保证维护的及时性,易造成测量结果与实际产生偏差,引发纠纷甚至安全事故。实施对计量设备状态参数的实时监测,不仅可提高计量设备的管理效率,而且可最大限度地减少安全隐患与经济损失^[1]。

天然气主要成分为烷烃,同时掺杂着许多其他气体成分,在对其种类和浓度进行监测的过程中不可避免地存在交叉敏感的问题,从而造成计量结构产生较大误差,故研究工作可从混合气体种类和浓度辨识 2 个层次上进行。目前最常用的辨识方法是气相色谱法,采用实时色谱仪实时取样分析,得到天然气组成,并可以在此基础上获得密度、发热量等物性参数,通过计算机进行数据实时记录,确保天然气流量计量准确可靠,保证后续气体的正常使用^[2-3]。气相色谱法的最大缺陷在于数据获得周期长,需要消耗大量载气,无法真正首先和实时监测。对此,本文提出采用红外气体传感器阵列结合信息融合技术辅助,实现天然气中混合气体组分的种类识别和浓度监测。首先利用开源的样本量较大的 UCI(University of California Irvine)数据库进行 6 种单一组分气体种类辨识继而进行浓度回归预测,根据准确度优选最佳算法方案,并将其用于实验室模拟烷烃中的气态组分甲烷、乙烷、丙烷以及用于增加压力而注入的氮气的小样本混合气体浓度预测,希望能够通过借助算法的手段,对气体浓度的偏差进行修正,结合红外传感阵列与可编程逻辑器件,实现真正的天然气气体浓度在线监测^[4-5]。

1 各种算法比较与筛选

测试使用的 UCI 数据库是加州大学欧文分校

的机器学习数据库。本文使用的数据集是其中关于数据回归类的 1 个,包含 13 910 个数据样本,由 16 个传感器对 6 种纯净单一组分在不同浓度下测得。每个传感器均记录 8 个特征值,对应每个样本的 128 个特征参数^[6]。

分别采用最近邻(KNN)算法、分类回归树(CART)算法以及多层感知机(MLP)算法对 6 种气体的类别进行了辨识,然后,分别将 3 种算法的预测结果进行了比较。该数据集中所测量的 6 种组分依次为氨气、乙醛、丙酮、乙烯、乙醇和甲苯,所测得的浓度范围如表 1 所示。

表 1 气体种类及浓度范围

组分	体积分数/ 10^{-6}
氨气	50~1 000
乙醛	5~500
丙酮	12~1 000
乙烯	10~300
乙醇	10~600
甲苯	10~100

具体的研究步骤如下:首先,使用主成分分析(PCA)法将数据特征降维,提取其主要特征,然后采用分类算法对 6 种组分进行分类,以分类结果准确率为指标判断算法的优劣。

1.1 气体种类辨识算法比较

1.1.1 PCA 降维 原数据为 128 维,使用 PCA 降维并提取贡献值之和超过 99.9%的前 n 个特征,分别用 P_{C1} 、 P_{C2} 和 P_{C3} 表示 1 维、2 维、3 维特征值,图 1 是提取前 3 个主成分特征的可视化结果,表明降维后气体可被分辨,其方差占比分别为 90.09%、6.08%与 2.49%。根据计算结果,当 $n=6$ 时,数据

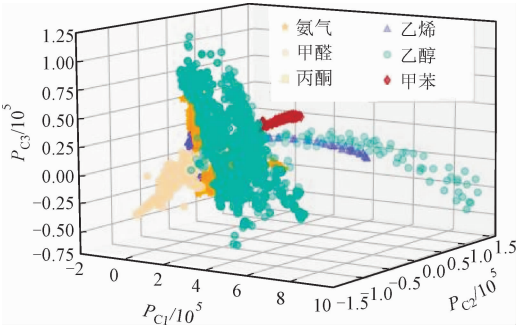


图 1 $n=3$ 时 PCA 特征降维后的可视化结果

特征已经达到 99.9%，可以进行种类辨别，故 KNN 和 CART 算法均在此基础上进行。在进行维度压缩之后，能够大大简化分类的数据量，同时保留原始数据最主要得特征，消除特征之间的线性相关，加快训练时间^[7]。

1.1.2 KNN、CART、MLP 算法分析 对于 KNN 算法，决定模型优劣的参数有 2 个：近邻个数和距离度量参数 p ($p=1$ 时为绝对距离， $p=2$ 为欧几里得距离)，将数据集按照 7:3 的比例分为训练集和测试集，在训练过程中采用了一折交叉验证法^[8-10]。图 2 为 $p=1, 2$ 时，所有样本的训练集、验证集、测试集与总集的辨识准确率随近邻数 k 的变化趋势。从图 2 中可以看出， $k=1$ 且 $p=1$ 时，KNN 算法对 6 种组分的分类效果达到最佳。由于所测试的组分的类型是确定的，因此可以根据每种组分的测试，使用同样方法设定不同的最优测试参数。

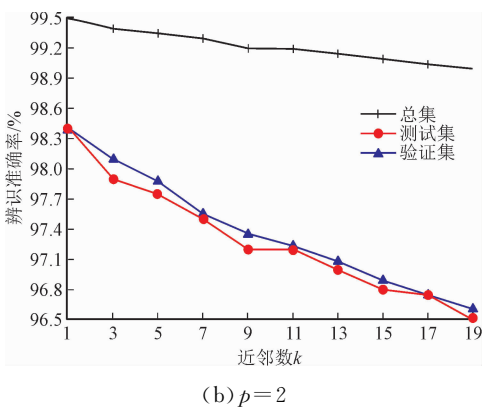
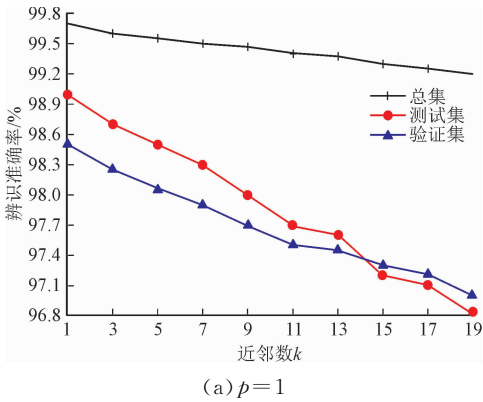


图 2 两种 p 参数下 6 组分的总辨识准确率随近邻数 k 的变化情况

CART 算法的分类效果与算法选择的终止条件有关。本文选择的终止条件为使每一个叶子结点的样本都为同一类，依然按照 6 个特征值和 7:3 的比例划分数据集^[11-12]。

MLP 神经网络的映射方式为主动映射，无需事前揭示描述这种映射关系的数学方程，它的学习规

则是使用梯度下降法。数据样本量越大，训练效果越好，所以训练时将 128 维特征全部输入^[13-15]。使用多层神经网络进行预测的时候，需要在每一层都尽可能多的提取到数据的特征，因而需要对数据进行预处理。这里分别采用 Min-Max Adam(MMA)、Min-Max RmsProp (MMR)、Min-Max Momentum (MMM)、Min-Max SGD (MMS)、Z-Score Adam (ZSA)、Z-Score RmsProp (ZSR)、Z-Score Momentum (ZSM) 和 Z-Score SGD (ZSS) 这 8 种方法对包含 1 个隐含层的神经网络进行训练，根据图 3 可知，最优的组合为 ZSA 算法。表 2 显示了 3 种算法得到的分类效果。

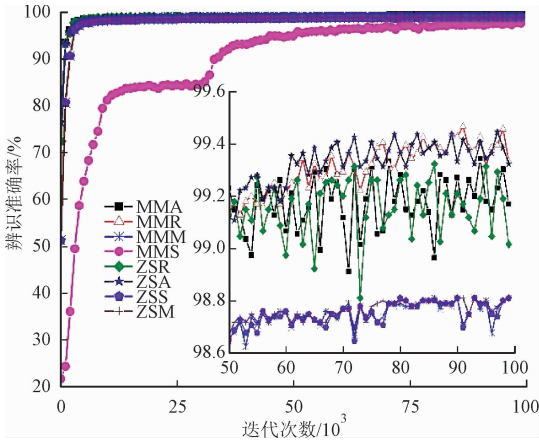


图 3 不同数据预处理方法的辨识准确率

表 2 3 种算法分类准确率比较

组分	准确率/%		
	KNN	CART	MLP
氨气	99.8	97.2	98.9
乙醛	99.6	98.2	99.1
丙酮	98.4	96.2	98.9
乙烯	99.3	95.9	98.2
乙醇	97.9	96.6	98.6
甲苯	98.7	96.9	99.0
总准确率/%	99.0	96.8	98.8

1.1.3 分类算法小结 3 种算法中，KNN 与 MLP 算法均在测试集上获得了超过 98% 的辨识准确率，而 CART 算法的气体辨识准确率只有约 96%。这是因为，CART 算法在每次决策时，只考虑 1 个维度的特征，而不考虑维度之间的关联度。此结论目前只在所选的样本实例中进行了验证，理论上可应用于同类型其他样本。

1.2 UCI 组分浓度预测

针对已经判定好类别的组分，使用回归算法，对

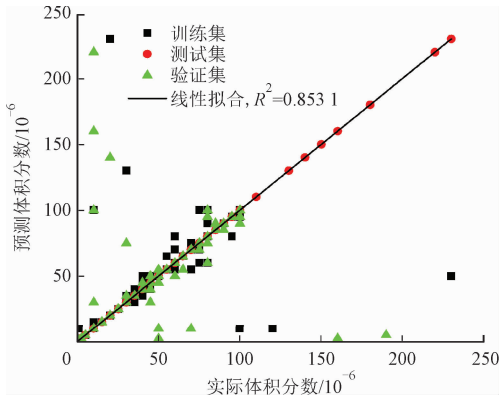
气体浓度进行预测,并采用最小二乘回归中的相关系数值来评估回归分析的好坏。分别使用了线性回归、KNN 回归、CART 回归和 MLP 回归算法对 6 种组分的浓度进行分析,给出最好与最差的预测效果图,并列出了以 R^2 为衡量指标的预测效果对比表。

1.2.1 几种回归算法分析 首先使用 PCA 算法对原始的 128 维数据降至 10 维,同时将原始样本集按照 7:3 的比例划分为训练集与测试集,分别绘制出 6 组分的真实浓度与预测浓度之间的关系,利用线性回归来判断数据之间是否有相关性,从而说明能够利用算法进行回归预测。

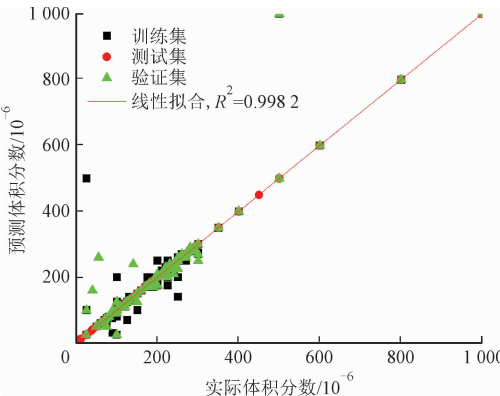
线性回归对 6 组分的预测效果相似,因此只需给出预测结果中线性相关程度最高和最低的,分别为乙烯与甲苯,如图 4a、图 4b 所示。预测结果说明所提取的特征与浓度之间有一定相关性,但不是简单地线性叠加关系,而是包含了更高次的关系,因此本文使用了其他方法对组分浓度进行预测。

本文使用 KNN 算法进行回归预测,将原始样本集按照 7:3 的比例划分为训练集与测试集。图 5 显示了使用 KNN 算法对甲苯和丙酮的预测效果,其 R^2 值分别为 0.998 2 和 0.853 1,其余的 4 种气体

也均达到了 0.97 以上,获得了比较好的预测结果。KNN 算法的最大优点是对异常值不敏感,缺点是样本不平衡问题(即有些类别的样本数量很多,而其他样本的数量很少),无法给出数据的内在含义,需要记录整个样本数据,硬件实现困难。



(a) 甲苯

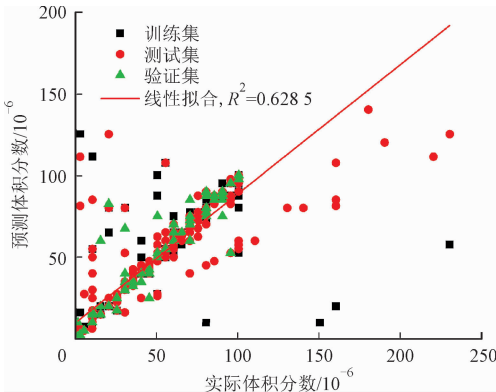


(b) 丙酮

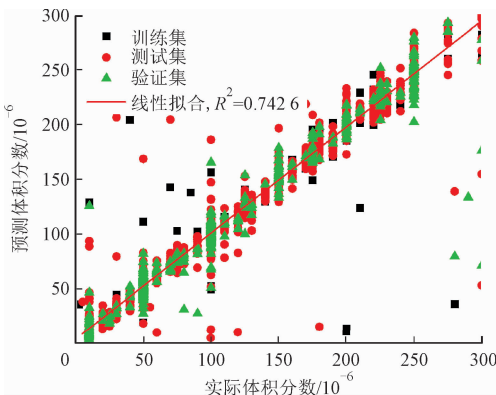
图 5 使用 KNN 算法的预测结果

KNN 算法通过比较样本距离进行回归,得到的预测结果较好,因此可以尝试使用与 KNN 回归思想相似的 CART 回归算法,其基本思想是将样本空间细分成若干子间,子空间内样本的输出 y (连续值) 的均值即为该子空间内的预测值。图 6 为 CART 算法的预测结果,由图 6 可以见: CART 算法对丙酮的预测效果最佳,其 R^2 值达到 0.986 0,对甲苯的预测效果最差,只有 0.838 4;整体预测效果好于线性,低于 KNN 算法。原因在于每次决策仅仅考虑了 1 个属性对预测值的影响,而不考虑属性之间隐含的关联关系,故预测效果会略差,并且生成的树过为庞大的话,不利于后续规则的提取,因而一般情况下可以选择将 CART 分类算法作为回归之前的一个预处理过程,在出现数据阶段性波动的情况下,先进行梯度判断分类,再进行阶段回归。

在尝试了上述 3 种方法之后,本文使用较经典

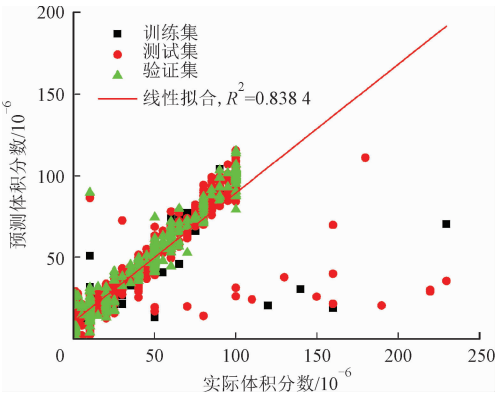


(a) 甲苯

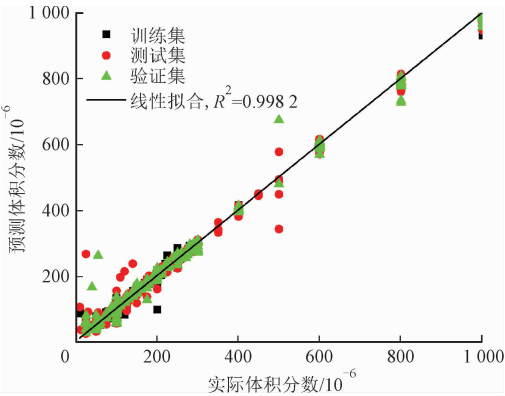


(b) 乙烯

图 4 使用线性回归算法的预测结果

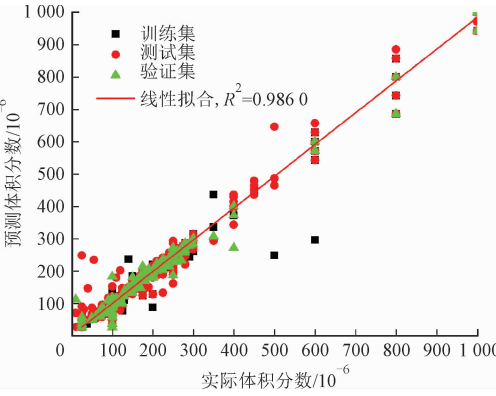


(a) 甲苯



(b) 丙酮

图 7 使用包含一个隐藏层的 MLP 网络的预测结果



(b) 丙酮

图 6 使用 CART 算法的预测结果

果较差,说明传感器的特征与气体浓度之间有高次依赖关系;其他 3 种算法中,KNN 回归与 MLP 回归在准确度上有较大优势,因而在对实验数据的分析上我们将选用这两种算法进行尝试。

表 3 4 种回归算法的精度对比

类别	相关系数 R^2			
	线性回归	KNN	CART	MLP
乙醇	0.613 4	0.975 5	0.933 9	0.976 9
乙烯	0.742 6	0.973 0	0.958 1	0.963 9
氨气	0.671 2	0.985 7	0.977 7	0.983 9
乙醛	0.721 4	0.990 4	0.963 8	0.969 7
丙酮	0.882 0	0.998 2	0.986 0	0.998 2
甲苯	0.628 5	0.853 1	0.838 4	0.900 1

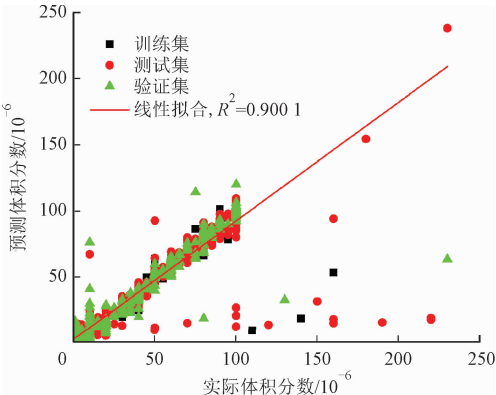
2 天然气数据浓度回归算法分析比较

2.1 数据集介绍

实验数据总共 4 238 个,在实验腔室内 4 组分混合气体(甲烷、乙烷、丙烷、氮气)条件下获得 3 个传感器(1 号为甲烷传感器,2 号为丙烷传感器,3 号为不同型号的丙烷红外传感器,均来自 Premier 公司)的读数,设定甲烷的实际体积分数测量范围为 0%~100%,乙烷设定为 0%~25%,丙烷设定为 0%~10%。特征值为 3 个传感器的读数以及测试腔室内的温度和压力补偿值,输出为预测的甲烷、乙烷、丙烷之和以及氮气的体积分数。表 4 列出了混合气体部分样本数据。

表 4 混合气体部分样本数据

体积分数实际值/%				传感器响应值/%			温度	压力
甲烷	乙烷	丙烷	氮气	1 号	2 号	3 号	/℃	/kPa
0.0	5.4	1.0	93.6	0.0	5.9	6.4	29.6	100
10.5	5.1	9.1	75.3	9.6	37.9	43.5	27.8	100
95.0	1.0	3.2	0.8	103.3	52.4	63.4	41.7	12



(a) 甲苯

2.2 KNN 算法回归预测结果

采用 1.2 节的回归算法进行测试。图 8 为使用 KNN 算法对混合气体浓度的预测结果,图中部分数据点存在较大误差是由于实验过程中不可避免的人为操作,以及传感器读数存在一定幅度摆动所致。

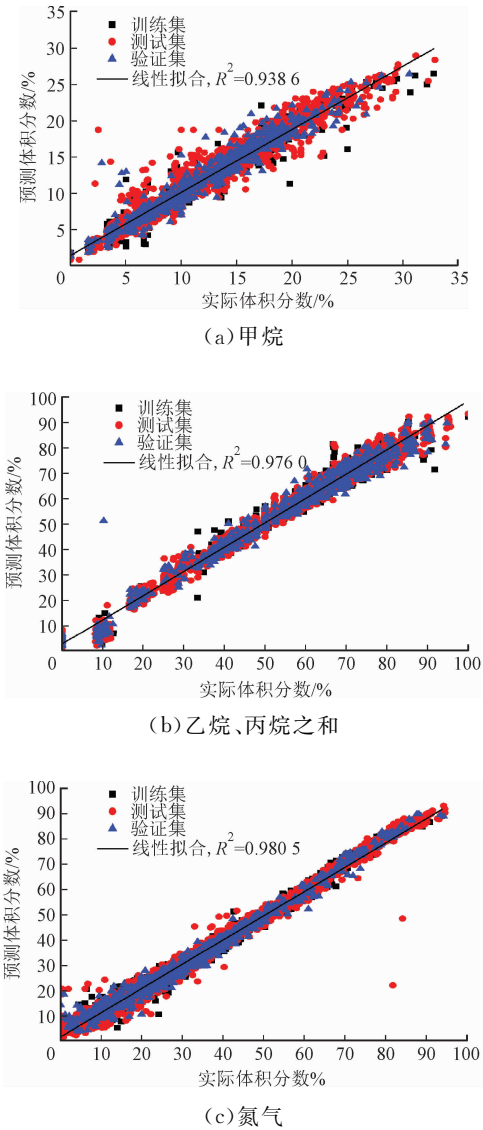


图 8 使用 KNN 算法对混合气体浓度预测结果

2.3 MLP 算法回归预测结果

与 2.2 节处理方法类似,神经网络结构为 5 输入(3 个红外传感器读数,温度与压力补偿)3 输出(甲烷,乙烷、丙烷之和及氮气的体积分数)。根据具体数据情况,经比较,决定采用 Z-Score 归一化进行样本预处理,采用 MLP 优化算法,预测结果如图 9 所示。

综合考虑,使用 MLP 算法的预测结果更佳,MLP 回归算法的预测准确性高,可调参数多,优化空间大。作为电子鼻设备的优化方案,只需要记录

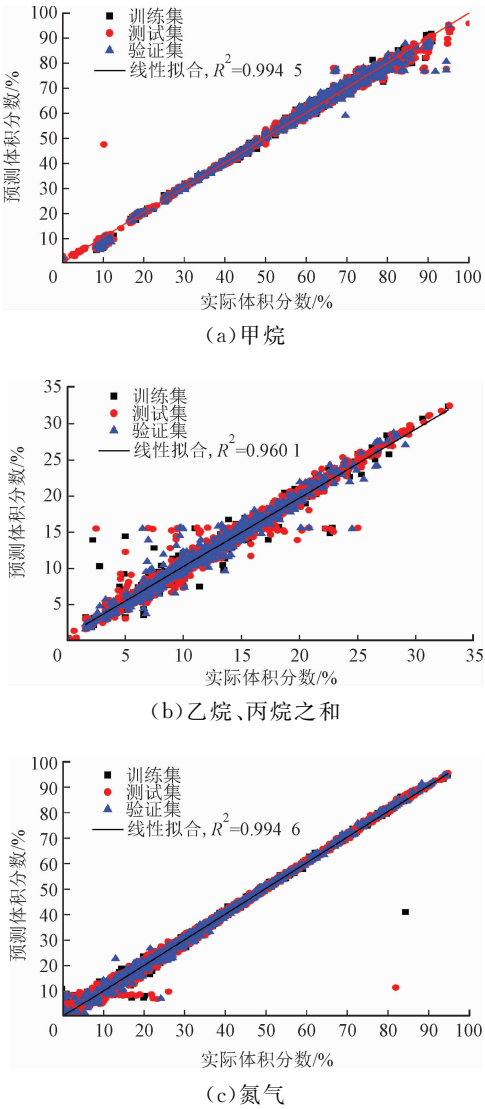


图 9 使用 MLP 算法对混合气体浓度预测结果

训练模型的权值与偏置,硬件可实现。

3 结 论

本文采用 KNN、CART 和 MLP 这 3 种分类算法对 UCI 数据样本库中的 6 种单一组分进行了分类预测,结果表明,KNN 与 MLP 算法对类别辨识的准确率均达到了 98% 以上,CART 算法的准确率为 96.98%。对不同种类的气体数据进行了线性回归、KNN 回归、CART 回归以及 MLP 回归网络 4 种算法进行筛选,并采用相关系数 R^2 对回归效果做出评估。结果发现,KNN 算法对除甲苯以外的气体的浓度辨识效果均最好,而 MLP 网络对甲苯的辨识效果最佳。将筛选得到的 KNN 与 MLP 回归算法用于实验室样本,也达到了较好的回归预测效果,其中使用 MLP 算法对甲烷、非甲烷总烃和氮气进行预测,拟合得到的回归系数的 R^2 分别达到

0.99、0.96 和 0.99。最终得出结论,KNN 算法和 MLP 算法对天然气以及 VOCs 数据集的分类效果较好,MLP 算法在浓度回归上有较大优势。将 KNN 算法所得最优相关模型(KNN 的整个训练样本及分类规则)或相关参数(MLP 的偏置与权重)利用可编程硬件进行存储,便可在在线监测设备中使用。

参考文献:

- [1] 徐阶,陈广仁,钱济人. 天然气计量设备实时监测系统[J]. 煤气与热力, 2014, 34(1): 34-36.
XU Jie, CHEN Guangren, QIAN Jiren. Real-time monitoring system for natural gas metering equipment[J]. Gas and Heat, 2014, 34(1): 34-36.
- [2] 青青,陈福权,张玮健. 用气相色谱仪快速分析管道天然气成分[J]. 化学分析计量, 2010(4): 85-87.
QING Qing, CHEN Fuquan, ZHANG Weijian. The composition of natural gas in pipeline was analyzed by gas chromatograph[J]. Chemistry, 2010(4): 85-87.
- [3] 张俊有,章祝云,赵玉斌. 可调谐二极管激光吸收光谱技术在甲烷气体浓度检测中的应用[J]. 科技风, 2014(5): 76-76.
ZHANG Junyou, ZHANG Zhuyun, ZHAO Yubin. Application of tunable diode laser absorption spectroscopy in methane concentration detection[J]. Technology Wind, 2014(5): 76-76.
- [4] JIA Tanghao, GUO Tianle, WANG Xuming, et al. Mixed natural gas online recognition device based on a neural network algorithm implemented by a FPGA[J]. Sensors, 2019, 19(9): 2090.
- [5] 耿志广,王希武,王寅龙,等. 基于人工神经网络的电子鼻对混合气体检测研究[J]. 现代计算机, 2010(5): 45-48.
GENG Zhiguang, WANG Xiwu, WANG Yinlong, et al. Study on detection of mixed gas by electronic nose based on artificial neural network[J]. Modern Computer, 2010(5): 45-48.
- [6] VERGARA A, VEMBU S, AYHAN T, et al. Chemical gas sensor drift compensation using classifier ensembles[J]. Sensors and Actuators; B Chemical, 2012, 166/167: 320-329.
- [7] FALEH R, OTHMAN M, KACHOURI A, et al. Recognition of O₃ concentration using WO₃ gas sensor and principal component analysis[C]//2014 1st International Conference on Advanced Technologies for Signal & Image Processing. Piscataway, NJ, USA: IEEE, 2014: 322-327.
- [8] 郁春江. 基于特征选择的 KNN 算法在煤炭勘查工作中的应用[J]. 煤炭技术, 2013(12): 130-131.
YU Chunjiang. KNN algorithm based on feature selection of application in coal exploration[J]. Coal Technology, 2013(12): 130-131.
- [9] YANG Jingli, SUN Zhen, CHEN Yinsheng. Fault detection using the clustering-KNN rule for gas sensor arrays[J]. Sensors, 2016, 16(12): 2069.
- [10] ADHIKARI S, SAHA S. Multiple classifier combination technique for sensor drift compensation using ANN & KNN[C]//2014 4th IEEE International Advance Computing Conference. Piscataway, NJ, USA: IEEE, 2014: 1184-1189.
- [11] HUWEI J. Hybrid approach of selecting hyperparameters of support vector machine for regression[J]. IEEE Transactions on Systems, 2006, 36(3): 699-709.
- [12] LI Han, SHI Xiajing, WENDY W, et al. Nanoparticle-structured sensing array materials and pattern recognition for VOC detection[J]. Sensors and Actuators; B Chemical, 2005, 106(1): 431-441.
- [13] SUN Yulong, LUO Dehan, LI Hui, et al. Detecting and identifying industrial gases by a method based on olfactory machine at different concentrations[J]. Journal of Electrical and Computer Engineering, 2018, 2018: 1092718.
- [14] LOUTFI A, CORADESCHI S, MANI G K, et al. Electronic noses for food quality: a review[J]. Journal of Food Engineering, 2015, 144: 103-111.
- [15] XU Liyuan, HE Jie, DUAN Shihong, et al. Comparison of machine learning algorithms for concentration detection and prediction of formaldehyde based on electronic nose[J]. Sensor Review, 2016, 36(2): 207-216.
- [16] ZHANG Shunping, XIE Changsheng, HU Mulin, et al. An entire feature extraction method of metal oxide gas sensors[J]. Sensors and Actuators; B Chemical, 2008, 132(1): 81-89.
- [17] LIU Jiao, LIU Jinfu, YU Daren, et al. Fault detection for gas turbine hot components based on a convolutional neural network[J]. Energies, 2018, 11(8): 2149.
- [18] KHALDI S, DIBI Z. Neural network technique for electronic nose based on high sensitivity sensors array[J]. Sensing and Imaging, 2019, 20(1): 15.
- [19] SUNDGREN H, WINQUIST F, LUKKARI I, et al. Artificial neural networks and gas sensor arrays: quantification of individual components in a gas mixture[J]. Measurement Science and Technology, 1991, 2(5): 464-469.

(编辑 刘杨)